On asymptotically efficient simulation of large deviation probabilities

A. B. Dieker and M. Mandjes
CWI
P.O. Box 94079
1090 GB Amsterdam, the Netherlands
and
University of Twente
Faculty of Mathematical Sciences
P.O. Box 217
7500 AE Enschede, the Netherlands

ABSTRACT

Let $\{\nu_\epsilon : \epsilon > 0\}$ be a family of probabilities for which the decay is governed by a large deviation principle, and consider the simulation of $\nu_{\epsilon_0}(A)$ for some fixed measurable set A and some $\epsilon_0 > 0$. We investigate the circumstances under which an exponentially twisted importance sampling distribution yields an asymptotically efficient estimator. Varadhan's lemma yields necessary and sufficient conditions, and these are shown to improve the conditions of Sadowsky [25]. This is illustrated by an example for which Sadowsky's conditions do not apply, while an efficient twist exists.

2000 Mathematics Subject Classification: primary 65C05; secondary 60F10, 60K10.
Keywords and Phrases: Asymptotic efficiency, Gaussian processes, importance sampling, large deviations
Note: The research was supported by the Netherlands Organization for Scientific Research (NWO) under grant 631.000.002.

1. Introduction

Given a probability distribution ν, we are interested in estimating a rare event probability $\nu(A)$. In direct Monte Carlo methods, the usual estimator is the proportion of times that A occurs in a certain number of independent samples from ν. However, an inherent problem of this approach is that many samples are needed to obtain a reliable estimate for $\nu(A)$. In fact, the required simulation time for estimating $\nu(A)$ may exceed any reasonable limit.

As an important special case, direct Monte Carlo methods are inappropriate for simulating large deviation probabilities. A family of probability measures $\{\nu_\epsilon : \epsilon > 0\}$ is said to satisfy a *large deviation principle* (LDP) if $\nu_\epsilon(A)$ decays exponentially as $\epsilon \to 0$ for a wide class of sets A. Given such a family, we refer to a probability of the form $\nu_{\epsilon_0}(A)$ for some $\epsilon_0 > 0$ and some event A as a *large deviation probability*. Probabilities of this type are encountered in many fields, e.g., statistics, operations research, information theory, and financial mathematics.

A widely used technique to estimate rare-event probabilities is *importance sampling*. In importance sampling, one samples from a probability measure λ different from ν_{ϵ_0}, such that the ν_{ϵ_0}-rare event becomes λ-likely. Often, one chooses a so-called *exponentially twisted* distribution for λ, but within this class there is still freedom to select a specific twisted distribution. To evaluate the changes of measure, efficiency criteria have been developed. In this paper, we use the *asymptotic efficiency* criterion.

Research initiated by the seminal paper of Siegmund [27] has shown that exponentially twisting is asymptotically efficient in specific cases. We mention in particular the estimation of the 'level-crossing' probability $P(X_1+\ldots+X_n > M,\ \text{for some } n)$ for real-valued i.i.d. random variables $X_1, X_2, \ldots$; see Lehtonen and Nyrhinen [23], who study the regime $M \to \infty$. Related results in a more general Markovian setting are obtained by Asmussen [1] and Lehtonen and Nyrhinen [22]. Collamore [8] extends this to a multi-dimensional setting.

Another example for which an exponential twist is known to yield asymptotic efficiency relates to the 'Cramér-type' probability $P(X_1 + \ldots + X_n \geq \gamma n)$ for $\gamma > \mathbb{E}X_1$, where $n \to \infty$. In case $X_1, X_2, \ldots$ have a special Markovian structure, this was found by Bucklew et al. [6]. Sadowsky [24] focuses on stability issues in the special case of i.i.d. random variables. To estimate $P(S_n \geq \gamma n)$ for generally distributed S_n, Sadowsky and Bucklew [26] show that that there exists an asymptotically efficient exponential twist if the Gärtner-Ellis theorem applies to $\{S_n/n\}$; this is also observed by Szechtman and Glynn [28].

However, it was noted that a successful application of an importance sampling distribution based on large deviation theory critically depends on the specific problem at hand. Glasserman and Wang [19] give variations on both the level-crossing problem and the Cramér-type problem, and show that exponential twists can be inefficient if the rare event A is irregular. In fact, they obtain the stronger result that the so-called relative error can even become unbounded in these examples. Similar observations have been made by Glasserman and Kou [17] in a queueing context.

Given the examples of efficient and inefficient simulation with exponentially twisted importance sampling distributions, it is natural to ask whether there exist necessary and sufficient conditions for asymptotic efficiency. In case the Gärtner-Ellis theorem applies, this question is studied by Sadowsky and Bucklew [26], while Sadowsky [25] extends these findings to a general abstract large deviation setting.

The necessary and sufficient conditions presented here have two advantages over those in [25]. The first is that the proof is elementary; the conditions follow straightforwardly from an application of Varadhan's Integral lemma. Therefore, we refer to these conditions as *Varadhan conditions*. Notice that this elementary lemma has already been applied earlier to derive efficiency properties of certain rare event estimators (see Glasserman et al. [16], Dupuis and Wang [14], and Glasserman and Li [18]).

The main result of this paper is that the elementary Varadhan conditions are shown to improve the conditions of Sadowsky [25]. To explain the improvements, it is important to realize that each set of *conditions* (the Varadhan conditions and Sadowsky's conditions) applies only under certain *assumptions*. On the one hand, the assumptions underlying the Varadhan conditions are less restrictive than Sadowsky's, so that the conditions apply in more situations. Notably, convexity of the large deviation rate function is not required. On the other hand, the Varadhan conditions themselves are better than Sadowsky's conditions, i.e., the Varadhan sufficiency condition is implied by Sadowsky's sufficiency condition, and vice versa for the necessary condition.

The use of the Varadhan conditions is illustrated by an example for which Sadowsky's results cannot be applied. In this example, asymptotically efficient simulation is possible, while a different approach, which seems perhaps more natural, turns out to be slower.

The paper is organized as follows. After providing the necessary preliminaries in Section 2, we state the Varadhan conditions in Section 3. Section 4 shows that these conditions improve those of Sadowsky. An example is worked out in Section 5, and Section 6 relates the use of a single exponential twist to other approaches.

2. Preliminaries

This section provides the basic background on importance sampling and asymptotic efficiency, and discusses their relationship with large deviation techniques. For a more detailed discussion on importance sampling and asymptotic efficiency, see Asmussen and Rubinstein [3], Heidelberger [21], and references therein. Valuable sources for large deviation techniques are the books by Dembo and Zeitouni [9] and Deuschel and Stroock [10].

2.1 Importance sampling

Let $\mathcal{X}$ be a topological space, equipped with some σ-field $\mathcal{B}$ containing the Borel σ-field. Given a probability measure ν on $(\mathcal{X}, \mathcal{B})$, we are interested in the simulation of the ν-probability of a given event $A \in \mathcal{B}$, where $\nu(A)$ is small. The idea of importance sampling is to sample from a different distribution on $(\mathcal{X}, \mathcal{B})$, say λ, for which A occurs more frequently. This is done by specifying a measurable function $d\lambda/d\nu : \mathcal{X} \to [0, \infty]$ and by setting

$$\lambda(B) := \int_B \frac{d\lambda}{d\nu} d\nu. \tag{2.1}$$

Since λ must be a probability measure, $d\lambda/d\nu$ should integrate to unity with respect to ν.

Assuming equivalence of the measures ν and λ, set $d\nu/d\lambda := (d\lambda/d\nu)^{-1}$ and note that

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} d\lambda = \int_{\mathcal{X}} \mathbf{1}_A \frac{d\nu}{d\lambda} d\lambda,$$

where $\mathbf{1}_A$ denotes the indicator function of A. The importance sampling estimator $\widehat{\nu_\lambda(A)}$ of $\nu(A)$ is found by drawing N independent samples $X^{(1)}, \ldots, X^{(N)}$ from λ:

$$\widehat{\nu_\lambda(A)} := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{X^{(i)} \in A\}} \frac{d\nu}{d\lambda}(X^{(i)}).$$

It is clear that $\widehat{\nu_\lambda(A)}$ is an unbiased estimator, i.e., $\mathbb{E}_\lambda \widehat{\nu_\lambda(A)} = \nu(A)$. However, one has the freedom to choose an efficient distribution λ in the sense that the variance of the estimator is small. In particular, it is of interest to find the change of measure that minimizes this variance, or, equivalently,

$$\int_A \left(\frac{d\nu}{d\lambda}\right)^2 d\lambda = \int_{\mathcal{X}} \mathbf{1}_A \left(\frac{d\nu}{d\lambda}\right)^2 d\lambda = \int_A \frac{d\nu}{d\lambda} d\nu.$$

A zero-variance estimator is found by letting λ be the conditional distribution of ν given A (see, e.g., Heidelberger [21]), but it is infeasible since $d\nu/d\lambda$ then depends on the *unknown* quantity $\nu(A)$. This motivates the use of another optimality criterion, *asymptotic efficiency*.

2.2 Asymptotic efficiency

To formalize the concept of asymptotic efficiency, we introduce some notions that are extensively used in large deviation theory.

A function $I : \mathcal{X} \to [0, \infty]$ is said to be lower semicontinuous if the level sets $\Phi_I(\alpha) := \{x : I(x) \leq \alpha\}$ are closed subsets of $\mathcal{X}$ for all $\alpha \in [0, \infty)$. The interior and closure of a set $B \subseteq \mathcal{X}$ are denoted by B^o and $\overline{B}$ respectively.

Definition 1 *A function $I : \mathcal{X} \to [0, \infty]$ is called a* rate function *if it is lower semicontinuous. If $\Phi_I(\alpha)$ is compact for every $\alpha \geq 0$, I is called a* good rate function.

*A set $B \in \mathcal{B}$ is called an I-*continuity set *if* $\inf_{x \in B^o} I(x) = \inf_{x \in B} I(x) = \inf_{x \in \overline{B}} I(x)$.

The central notion in large deviation theory is the large deviation principle.

Definition 2 *A family of probability measures $\{\nu_\epsilon : \epsilon > 0\}$ on $(\mathcal{X}, \mathcal{B})$ satisfies a large deviation principle (LDP) with rate function I if for all $B \in \mathcal{B}$,*

$$-\inf_{x \in B^o} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(B) \leq \limsup_{\epsilon \to 0} \epsilon \log \nu_\epsilon(B) \leq -\inf_{x \in \overline{B}} I(x).$$

Throughout this paper, we assume that the family $\{\nu_\epsilon\}$ satisfies an LDP. We fix some *rare event* $A \in \mathcal{B}$, i.e., $\inf_{x \in A^o} I(x) > 0$, implying that $\nu_\epsilon(A)$ decays exponentially as $\epsilon \to 0$. Since $\nu(A)$ is supposed to be a large deviation probability, we have $\nu = \nu_{\epsilon_0}$ for some ϵ_0.

The definition of asymptotic efficiency is related to the so-called *relative error*. Consider an i.i.d. sample $X^{(1)}_{\lambda_\epsilon}, \ldots, X^{(N)}_{\lambda_\epsilon}$ from an importance sampling distribution λ_ϵ. We define the relative error $\eta_N(\lambda_\epsilon, A)$ of the importance sampling estimator

$$\widehat{\nu_{\lambda_\epsilon}(A)}_N := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\left\{X^{(i)}_{\lambda_\epsilon} \in A\right\}} \frac{d\nu_\epsilon}{d\lambda_\epsilon} \left(X^{(i)}_{\lambda_\epsilon} \right) \tag{2.2}$$

by

$$\eta_N^2(\lambda_\epsilon, A) := \frac{\mathbb{V}\mathrm{ar}_{\lambda_\epsilon} \widehat{\nu_{\lambda_\epsilon}(A)}_N}{\nu_\epsilon(A)^2} = \frac{\mathbb{E}_{\lambda_\epsilon} \left(\widehat{\nu_{\lambda_\epsilon}(A)}_N \right)^2}{\nu_\epsilon(A)^2} - 1.$$

The relative error is proportional to the width of a confidence interval relative to the (expected) estimate itself; hence, it measures the variability of $\widehat{\nu_{\lambda_\epsilon}(A)}_N$.

For asymptotic efficiency, the number of samples required to obtain a prespecified relative error should vanish on an exponential scale. Set $N^*_{\lambda_\epsilon} := \inf\{N \in \mathbb{N} : \eta_N(\lambda_\epsilon, A) \leq \eta_{\max}\}$.

Definition 3 *An importance sampling family $\{\lambda_\epsilon\}$ is called* asymptotically efficient *if*

$$\limsup_{\epsilon \to 0} \epsilon \log N^*_{\lambda_\epsilon} = 0, \tag{2.3}$$

for some given maximal relative error $0 < \eta_{\max} < \infty$.

In the literature, asymptotic efficiency is sometimes referred to as *asymptotic optimality*, *logarithmic efficiency*, or *weak efficiency*. We will shortly see that the specific value of $\eta_{\max}$ is irrelevant.

Let us briefly relate Definition 3 to other frequently used definitions for asymptotic efficiency. By definition of $\eta_N(\lambda_\epsilon, A)$, we have

$$\begin{aligned} N^*_{\lambda_\epsilon} &= \inf\left\{ N \in \mathbb{N} : \frac{1}{N}\int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon \le (\eta^2_{\max}+1)\nu_\epsilon(A)^2 \right\} \\ &= \left\lceil \frac{\int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon}{(\eta^2_{\max}+1)\nu_\epsilon(A)^2} \right\rceil . \end{aligned} \tag{2.4}$$

Equation (2.4) implies

$$\limsup_{\epsilon\to 0} \epsilon \log N^*_{\lambda_\epsilon} \le \limsup_{\epsilon\to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon - 2 \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A), \tag{2.5}$$

with equality if the limit $\lim_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A)$ exists. Sufficient for the existence of this limit is that A be an I-continuity set; in that case, $\lim_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A) = -\inf_{x\in A} I(x)$. In many applications, A is indeed an I-continuity set, in which case asymptotic efficiency is equivalent to

$$\limsup_{\epsilon\to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon \le -2 \inf_{x\in \overline{A}} I(x). \tag{2.6}$$

In turn, this is equivalent to $\lim_{\epsilon\to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon = -2 \inf_{x\in\overline{A}} I(x)$ by Jensen's inequality. By similar arguments, one can also readily see that

$$\limsup_{\epsilon\to 0} \frac{\log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon}{\log \nu_\epsilon(A)} \ge 2$$

is equivalent to asymptotic efficiency when A is an I-continuity set. Again, the corresponding lower bound follows from Jensen's inequality.

3. The Varadhan conditions for efficiency of exponential twisting

This section investigates the asymptotic efficiency of the estimators that are based on an exponential twist. After formalizing the imposed assumptions, we state necessary and sufficient conditions based on Varadhan's lemma. Section 4 discusses the relation with the conditions developed by Sadowsky [25].

Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ be a σ-field on $\mathcal{X}$ containing the Borel σ-field. We assume that $\mathcal{X}$ is also a vector space, but not necessarily a topological vector space. Throughout this section, we fix a rare event $A \in \mathcal{B}$ and a continuous linear functional $\xi : \mathcal{X} \to \mathbb{R}$. Having a topological vector space in mind, we write $\langle \xi, \cdot \rangle$ for $\xi(\cdot)$. We are given a family $\{\nu_\epsilon\}$ of probability measures on $(\mathcal{X}, \mathcal{B})$.

Assumption 1 (Varadhan assumptions) *Assume that*

(i) $\mathcal{X}$ *is a vector space endowed with some regular Hausdorff topology,*

(ii) $\{\nu_\epsilon\}$ *satisfies the LDP with a good rate function* I*,*

(iii) it holds that

$$\lim_{M\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\int_{\{x\in\mathcal{X}:\langle\xi,x\rangle\geq M\}}\exp[\langle\xi,x\rangle/\epsilon]\nu_\epsilon(dx)=-\infty, \tag{3.1}$$

and similarly for ξ replaced by $-\xi$,

We note that a simple sufficient condition for (3.1) to hold is given in Lemma 4.3.8 of [9]: $\limsup_{\epsilon\to 0}\epsilon\log\int\exp[\gamma\langle\xi,x\rangle/\epsilon]\nu_\epsilon(dx)<\infty$ for some $\gamma>1$; similarly for ξ replaced by $-\xi$.

A new family of probability measures $\{\lambda_\epsilon^\xi\}$ is defined by

$$\begin{aligned}\frac{d\lambda_\epsilon^\xi}{d\nu_\epsilon}(x) \;&:=\; \exp\left(\langle\xi,x\rangle/\epsilon-\log\int_{\mathcal{X}}\exp[\langle\xi,y\rangle/\epsilon]\nu_\epsilon(dy)\right) \\ &=\; \frac{\exp[\langle\xi,x\rangle/\epsilon]}{\int_{\mathcal{X}}\exp[\langle\xi,y\rangle/\epsilon]\nu_\epsilon(dy)}.\end{aligned} \tag{3.2}$$

The measures $\{\lambda_\epsilon^\xi\}$ are called *exponentially twisted with twist* ξ. If the family $\{\lambda_\epsilon^\xi\}$ is asymptotically efficient, we simply call the exponential twist ξ asymptotically efficient.

The following proposition plays a key role in the proofs of this section.

Proposition 1 *Let $d\lambda_\epsilon^\xi/d\nu_\epsilon$ be given by (3.2), and let $B\in\mathcal{B}$. Under Assumption 1, we have*

$$\begin{aligned}\liminf_{\epsilon\to 0}\epsilon\log\int_B\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi \;&\geq\; -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]-\inf_{x\in B^o}[I(x)+\langle\xi,x\rangle], \\ \limsup_{\epsilon\to 0}\epsilon\log\int_B\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi \;&\leq\; -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]-\inf_{x\in\overline{B}}[I(x)+\langle\xi,x\rangle].\end{aligned}$$

Proof. Fix $B\in\mathcal{B}$ and note that

$$\begin{aligned}&\epsilon\log\int_B\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi=\epsilon\log\int_B\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}d\nu_\epsilon \\ =\;&\epsilon\log\int_B\exp\left(\log\int_{\mathcal{X}}\exp(\langle\xi,y\rangle/\epsilon)\nu_\epsilon(dy)-\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx) \\ =\;&\epsilon\log\int_{\mathcal{X}}\exp\left(\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx)+\epsilon\log\int_B\exp\left(-\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx).\end{aligned} \tag{3.3}$$

By Assumption 1 and the continuity of the functional ξ, Varadhan's Integral Lemma (Theorem 4.3.1 in [9]) applies. Thus, the limit of the first term exists and equals

$$\lim_{\epsilon\to 0}\epsilon\log\int_{\mathcal{X}}\exp\left(\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx)=\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle-I(x)].$$

A similar argument can be applied to the second term in (3.3). The conditions of Varadhan's Integral Lemma are again satisfied to apply the lemma to the continuous functional $-\xi$. Now we use a variant of this lemma (see, e.g., Exercise 4.3.11 of [9]) to see that for any open set G and any closed set F

$$\begin{aligned}\liminf_{\epsilon\to 0}\epsilon\log\int_G\exp\left(-\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx) \;&\geq\; -\inf_{x\in G}[I(x)+\langle\xi,x\rangle], \\ \limsup_{\epsilon\to 0}\epsilon\log\int_F\exp\left(-\langle\xi,x\rangle/\epsilon\right)\nu_\epsilon(dx) \;&\leq\; -\inf_{x\in F}[I(x)+\langle\xi,x\rangle].\end{aligned}$$

In particular, these inequalities hold for B^o and $\overline{B}$. The claim follows by adding the two terms in (3.3) [using the fact that the limit of the first term exists]. □

The necessary and sufficient conditions, formulated in the next theorem, follow almost immediately from Proposition 1.

Theorem 1 (Varadhan conditions) *Let Assumption 1 hold. The exponential twist ξ is asymptotically efficient if*

$$\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]+\inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle]\geq 2\inf_{x\in A^o}I(x). \tag{3.4}$$

Let Assumption 1 hold and let A be an I-continuity set. If the exponential twist ξ is asymptotically efficient, then

$$\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]+\inf_{x\in A^o}[I(x)+\langle\xi,x\rangle]\geq 2\inf_{x\in\overline{A}}I(x). \tag{3.5}$$

Proof. Sufficiency follows from (2.5), the upper bound of Proposition 1, and the LDP of Assumption 1(ii):

$$\begin{aligned}\limsup_{\epsilon\to 0}\epsilon\log N^*_{\lambda^\xi_\epsilon} &\leq \limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda^\xi_\epsilon}\right)^2 d\lambda^\xi_\epsilon-2\liminf_{\epsilon\to 0}\epsilon\log\nu_\epsilon(A)\\ &\leq -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]-\inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle]+2\inf_{x\in A^o}I(x).\end{aligned}$$

For necessity the argument is similar. Note that the lower bound of Proposition 1 implies that

$$\limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda^\xi_\epsilon}\right)^2 d\lambda^\xi_\epsilon\geq -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]-\inf_{x\in A^o}[I(x)+\langle\xi,x\rangle].$$

Moreover, by the large deviation upper bound,

$$\liminf_{\epsilon\to 0}\epsilon\log\nu_\epsilon(A)\leq -\inf_{x\in\overline{A}}I(x).$$

Combining these observations with the assumption that A is an I-continuity set, we have

$$\begin{aligned}0 &= \limsup_{\epsilon\to 0}\epsilon\log N^*_{\lambda^\xi_\epsilon}\\ &= \limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda^\xi_\epsilon}\right)^2 d\lambda^\xi_\epsilon-2\lim_{\epsilon\to 0}\epsilon\log\nu_\epsilon(A)\\ &\geq -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]-\inf_{x\in A^o}[I(x)+\langle\xi,x\rangle]+2\inf_{x\in\overline{A}}I(x),\end{aligned}$$

as desired. □

As suggested by the form of Theorem 1, the sufficient condition is also necessary under a weak condition on the set A. We formalize this in the following corollary, which follows straightforwardly from Theorem 1.

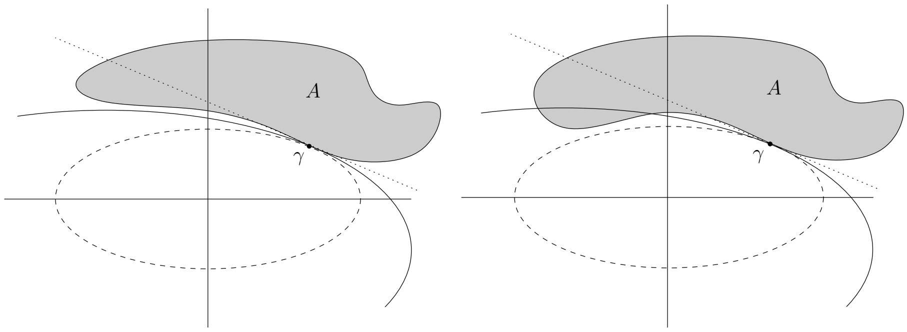

Figure 1: Efficient simulation with twist ξ_γ (left) and inefficient simulation with twist ξ_γ (right).

Corollary 1 *Let Assumption 1 hold, and assume that A is both an I-continuity set and an $(I+\xi)$-continuity set. Exponentially twisting with ξ is asymptotically efficient if and only if*

$$\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle]+\inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle]=2\inf_{x\in\overline{A}}I(x).$$

We remark that Sadowsky [25] uses a more general notion than asymptotic efficiency, namely ν-efficiency. Given an I-continuity set A, the importance sampling distribution λ_ϵ^ξ is said to be ν-efficient if

$$\limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^\nu d\lambda_\epsilon\leq -\nu\inf_{x\in\overline{A}}I(x).$$

In this terminology, we have established conditions for 2-efficiency (see the remarks after Definition 3). To obtain conditions for ν-efficiency with general $\nu\geq 2$, the statements in the subsection are easily modified. As an example, when A is an $(I+(\nu-1)\xi)$-continuity set and when Assumption 1(iii) holds with ξ replaced by $(\nu-1)\xi$ and $-(\nu-1)\xi$, the exponential twist ξ is ν-efficient if and only if

$$\inf_{x\in\mathcal{X}}[I(x)-(\nu-1)\langle\xi,x\rangle]+\inf_{x\in\overline{A}}[I(x)+(\nu-1)\langle\xi,x\rangle]=\nu\inf_{x\in\overline{A}}I(x).$$

We now illustrate the Varadhan conditions in a simple example. Let ν be the distribution of a random variable X on $\mathbb{R}^d$, and denote the distribution of the sample mean of n i.i.d. copies of X by ν_n. Note that $1/n$ plays the role of ϵ in this example. Let ν be such that Cramér's theorem holds.

For instance, ν_n is a zero-mean bivariate Gaussian distribution with covariance of the form Σ/n for some diagonal matrix Σ; see Figure 1. We are interested in $\nu_n(A)$ for two different sets A; these are drawn in the left and right panel of Figure 1. Note that the rate function has the form $I(x_1,x_2)=C_1x_1^2+C_2x_2^2$ for some constants $C_1,C_2>0$. As indicated by the dashed level curve of I, the 'most likely point' in A is in both cases γ, i.e., $\arg\inf_{x\in A}I(x)=I(\gamma)$. One can see that there is only one exponential twist $\xi_\gamma\in\mathbb{R}^2$ interesting for simulation purposes, namely the conjugate point of γ. The level curve of $I+\xi_\gamma+\inf_{x\in\mathcal{X}}[I(x)-\xi_\gamma^t x]$ that goes through γ is depicted as a solid line. Since both sets A are I- and $(I+\xi_\gamma)$-continuity sets, the twist ξ_γ is asymptotically efficient if and only if A lies entirely 'outside' the solid level curve (see Corollary 1). Hence, in the left panel the twist ξ_γ is asymptotically efficient twist and in the right panel it is not.

In the literature, sufficient conditions for asymptotic efficiency have been given in terms of dominating points and convexity of A in case the rate function is convex (see, e.g., Sadowsky and Bucklew [26] and references therein). Using Figure 1, we explain how it can be seen that the Varadhan conditions improve these dominating point conditions (convexity of A implies the existence of a dominating point, so we focus on dominating points). Every I-continuity set that touches γ and that is contained in the halfspace above the dotted line has dominating point γ. Obviously, such a set lies outside the solid level curve, and one can therefore estimate $\nu_n(A)$ asymptotically efficiently by an exponential twist. However, Figure 1 indicates that the dominating point condition is far from necessary: neither of the sets A have a dominating point, while an efficient twist exists in the left panel.

4. Comparison with Sadowsky's conditions

General necessary and sufficient conditions for asymptotic efficiency were also developed by Sadowsky [25]. In this subsection, we compare the conditions of Theorem 1 with Sadowsky's conditions. We show that the assumptions underlying Varadhan's conditions are less restrictive than Sadowsky's assumptions. Moreover, the sufficient condition in Theorem 1 improves Sadowsky's sufficiency condition, and the same holds for the accompanying necessary conditions.

In addition to the notation of the preceding subsection, we first introduce some new notions. In this section, $\mathcal{X}$ denotes a topological vector space, and $\mathcal{X}^*$ denotes the space of linear continuous functionals $\xi : \mathcal{X} \to \mathbb{R}$. Let $f : \mathcal{X} \to (-\infty, \infty]$ be a convex function. A point $x \in \mathcal{X}$ is called an *exposed point* of f if there exists a $\delta \in \mathcal{X}^*$ such that $f(y) > f(x) + \langle \delta, y - x\rangle$ for all $y \neq x$. δ is then called an *exposing hyperplane* of I at x.

To compare the Varadhan conditions to Sadowsky's, we first recapitulate Sadowsky's assumptions.

Assumption 2 (Sadowsky's assumptions) *Assume that*

(i) $\mathcal{X}$ *is a locally convex Hausdorff topological vector space,*

(ii) $\{\nu_\epsilon\}$ *satisfies the LDP with a* convex *good rate function* I*,*

(iii) for every $\delta \in \mathcal{X}^*$*,*

$$\Lambda(\delta) := \limsup_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} \exp[\langle \delta, x\rangle/\epsilon] \nu_\epsilon(dx) < \infty,$$

(iv) A *satisfies*

$$0 < \inf_{x \in A^o \cap \mathcal{F}} I(x) = \inf_{x \in A} I(x) = \inf_{x \in \overline{A}} I(x) < \infty,$$

where $\mathcal{F}$ *denotes the set of exposed points of* I*.*

Although Assumption 2 looks very similar to Assumption 1, there are crucial differences. To start with, $\mathcal{X}$ is not assumed to be a *topological* vector space in Assumption 1(i). To see the importance of this difference for applications, note that the space $D([0,1],\mathbb{R})$ of càdlàg functions on $[0,1]$ with values in $\mathbb{R}$ is a (regular, Hausdorff) vector space but no topological vector space when equipped with the Skorohod topology. We stress that the regularity of $\mathcal{X}$ assumed in Assumption 1(i) is implicit in Assumption 2(i): any real Hausdorff topological vector space is regular.

Moreover, the convexity of the large deviation rate function is not assumed in Assumption 1(ii). Note that this convexity is granted when an LDP is derived using an (abstract) Gärtner-Ellis type theorem, but non-convex rate functions also arise naturally in applications; see Section 5 for a discussion. Assumption 2(iii) implies Assumption 1 since $\gamma\xi$ is a continuous linear functional for any $\gamma \in \mathbb{R}$, while the fourth part of Assumption 2 is slightly stronger than requiring that A be an I-continuity set.

In the above comparison between Assumption 1 and Assumption 2, we have shown the following.

Proposition 2 *Assumption 2 implies that Assumption 1 holds and that A is an I-continuity set.*

In the remainder of this subsection, we compare the necessary and sufficient conditions of Theorem 1 to the conditions in [25]. Such a comparison is only possible when Sadowsky's assumption hold, i.e., we have to impose the (strong) Assumption 2. We start by repeating Sadowsky's conditions. Given that Assumption 2(iv) holds for A, we call $\gamma \in \overline{A}$ a *point of continuity* if $I(\gamma) = \inf_{x\in\overline{A}} I(x)$ and there exists a sequence $\{\gamma_n\} \subset A^o \cap \mathcal{F}$ such that $\gamma_n \to \gamma$.

Theorem 2 (Sadowsky's conditions) *Let Assumption 2 hold. The exponential twist ξ is asymptotically efficient if*

(a) there is a point of continuity γ such that $I(\gamma) = \langle \xi, \gamma\rangle - \Lambda(\xi)$,

(b) $I(x) + \langle \xi, x\rangle \geq I(\gamma) + \langle \xi, \gamma\rangle$ for all $x \in \overline{A}$,

(c) either $\langle \xi, x\rangle \geq \langle \xi, \gamma\rangle$ for all $x \in \overline{A}$, or there exists an $x \in \mathcal{F}$ such that ξ is an exposing hyperplane of I at x.

Let Assumption 2 hold. If the twist ξ is asymptotically efficient, then

(a) there is a point of continuity γ such that $I(\gamma) = \langle \xi, \gamma\rangle - \Lambda(\xi)$,

($\tilde{b}$) $I(x) + \langle \xi, x\rangle \geq I(\gamma) + \langle \xi, \gamma\rangle$ for all $x \in A^o \cap \mathcal{F}$.

Proposition 3 *Let Assumption 2 hold. The sufficient condition in Theorem 2 implies the sufficient condition in Theorem 1.*

Proof. By condition (a) of Theorem 2, there exists a point of continuity $\gamma \in \overline{A}$ such that $I(\gamma) = \inf_{x\in\overline{A}} I(x) = \langle \xi, \gamma\rangle - \Lambda(\xi)$. Since we assume that an LDP holds for some convex I [Assumption 2(ii)] and that Assumption 2(iii) holds, by Theorem 4.5.10(b) in [9] we have $I(x) = \sup_{\delta\in\mathcal{X}^*}[\langle \delta, x\rangle - \Lambda(\delta)]$, and hence $I(x) \geq \langle \xi, x\rangle - \Lambda(\xi)$. Combining this with $I(\gamma) = \langle \xi, \gamma\rangle - \Lambda(\xi)$, we conclude

$$\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] \geq -\Lambda(\xi) = I(\gamma) - \langle \xi, \gamma\rangle,$$

where the inequality may obviously be replaced by an equality.

It is immediate from condition (b) of Theorem 2 that $\inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle]=I(\gamma)+\langle\xi,\gamma\rangle$. Since $\inf_{x\in A^o} I(x)=I(\gamma)$, this implies the sufficient condition (3.4) in Theorem 1. □

It is important to notice that we did not use part (c) of Sadowsky's sufficient condition in the proof of Proposition 3; this part is redundant.

Proposition 4 *Let Assumption 2 hold. The necessary condition in Theorem 2 is implied by the necessary condition in Theorem 1.*

Proof. Let the twist ξ be asymptotically efficient. We start by showing that a point of continuity exists under Assumption 2. First note that $\inf_{x\in A^o\cap\mathcal{F}} I(x)=\inf_{x\in\overline{A}} I(x)$ [Assumption 2(iv)] implies that for any $n\in\mathbb{N}$ one can find some $\gamma_n\in A^o\cap\mathcal{F}\cap K_n$, where

$$K_n:=\{x\in\mathcal{X}:I(x)\le\inf_{y\in\overline{A}} I(y)+1/n\}.$$

Use $\inf_{x\in\overline{A}} I(x)<\infty$ and the goodness of the rate function [Assumption 2(ii)] to see that K_n is a compact subset of $\mathcal{X}$, hence also sequentially compact. Since K_n decreases in n, we obviously have $\{\gamma_n\}\subset K_1$. Hence, one can substract a subsequence that converges, say, to $\gamma\in K_1$. Since K_n is closed for every n and $\{\gamma_\cdot\}$ is eventually in K_n, we must also have that $\gamma\in K_n$ for every n. As a consequence, we have $I(\gamma)\le\inf_{x\in\overline{A}} I(x)$. Moreover, since $\{\gamma_n\}\subset A^o\cap\mathcal{F}$, we also see that $\gamma\in\overline{A^o\cap\mathcal{F}}\subset\overline{A}$. Therefore, $I(\gamma)=\inf_{x\in A^o\cap\mathcal{F}} I(x)=\inf_{x\in\overline{A}} I(x)$, and γ is a point of continuity.

The necessary condition in Theorem 1 implies

$$\begin{aligned}2I(\gamma) &\le \inf_{x\in A^o}[I(x)+\langle\xi,x\rangle]-\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle-I(x)]\\ &\le \lim_{n\to\infty}[I(\gamma_n)+\langle\xi,\gamma_n\rangle]-[\langle\xi,\gamma\rangle-I(\gamma)]=2I(\gamma).\end{aligned}$$

As a result, the inequalities can be replaced by equalities and we obtain

$$\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle-I(x)]=\langle\xi,\gamma\rangle-I(\gamma)\ \text{ and }\ \inf_{x\in A^o}[I(x)+\langle\xi,x\rangle]=I(\gamma)+\langle\xi,\gamma\rangle.$$

By Theorem 4.5.10(a) of [9], we also have $\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle-I(x)]=\Lambda(\xi)$ under Assumption 2. Hence, $I(\gamma)=\langle\xi,\gamma\rangle-\Lambda(\xi)$ and part (a) of Sadowsky's necessary condition is derived. Part $(\tilde{\text{b}})$ is immediate by noting that $\inf_{x\in A^o}[I(x)+\langle\xi,x\rangle]=I(\gamma)+\langle\xi,\gamma\rangle$ implies that $I(x)+\langle\xi,x\rangle\ge I(\gamma)+\langle\xi,\gamma\rangle$ for all $x\in A^o$. □

5. An example

In this section, we provide an example showing how Corollary 1 is typically used. The conditions of Sadowsky [25] do not apply to this example, since the rate function is non-convex. Despite this non-convexity, we show that an exponential twist may still be asymptotically efficient.

Non-convex rate functions arise naturally in several large deviation settings. Notably, certain large deviations of Markov chains and Markov processes (such as diffusions) are governed by rate functions that need not be convex; see, e.g., Feng and Kurtz [15]. Intuitively, analyzing the rate of convergence of random functions to a non-linear (deterministic) function causes the rate function to be non-convex. Non-convex rate functions also appear when investigating the rate of convergence to a non-degenerate measure, and the example of this section is of the latter type.

For our example, it appears that the event under consideration can be cut into disjoint 'subevents', that comply with Sadowsky's conditions. However, such a 'cut approach' might be impossible in other cases, or lead to a large number of subevents (that need to be estimated separately). This is especially relevant for the simulation of hitting probabilities of stochastic processes, as in Collamore [8]. Depending on the shape of the hitting curve, the simulation with a single twist may work, but the 'cut approach' leads to simulation of many events (each of which corresponds to hitting the curve at a particular point in time).

To avoid technicalities that are irrelevant for this paper, we do not illustrate the Varadhan conditions in a sample path setting. Instead, we discuss a relatively simple example, which gives at the same time a good feel under what circumstances the Varadhan conditions are more useful than Sadowsky's conditions.

Let us first give some background on our example. Recall that a phase-type distribution is a distribution associated to a finite Markov process, which can be characterized by three quantities $(E, \boldsymbol{\alpha}, \boldsymbol{T})$, see, e.g., Asmussen [2, Ch. III.4]. Given an arbitrary distribution ν on $(0,\infty)$, one can find a sequence of phase-type distributions that converges weakly to ν [2, Thm. III.4.2]. This implies that $\nu_n(A) \to 0$ for a large number of sets A; in fact, $\nu_n(A)$ then vanishes at an exponential rate. We are interested in $\nu_n(A)$ for fixed n.

We consider a particularly simple distribution, namely one that is concentrated on $\{1,5\}$. We write $\alpha = \nu(\{1\}) = 1 - \nu(\{5\})$. It is notationally cumbersome to describe a sequence of phase-type distributions that converges weakly to ν in the $(E, \boldsymbol{\alpha}, \boldsymbol{T})$-notation; a direct description is more appropriate here. Define ν_n as the distribution of Y_n, where Y_n has an Erlang(n, n) distribution with probability α and an Erlang(n, $n/5$) distribution with probability $1-\alpha$. (Recall that the sum of k independent exponentially distributed random variables with parameter λ has an Erlang(k, λ) distribution.) It is left to the reader to check that ν_n converges weakly to ν.

The approximating phase-type distributions are special cases of *mixtures*, and the phenomena that we observe in our example are typically also encountered in the mixture setting. Indeed, the large deviations of some mixtures are governed by non-convex rate functions. This holds in particular for finite mixtures (in our case, ν_n is a mixture of two distributions), for which the large deviations are readily analyzed. The infinite case is non-trivial; see Dinwoodie and Zabell [12, 13], Chaganty [7], and Biggins [4]. Importantly, mixtures also arise naturally in connection with conditional probabilities, see, e.g., [18].

Consider the simulation of $\nu_n(A)$, where $A := [0, 1/10) \cup (10, \infty)$. It is easy to see that ν_n equals the distribution of

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n X_i & \text{with probability } \alpha; \\ \frac{5}{n}\sum_{i=1}^n X_i & \text{with probability } 1-\alpha. \end{cases}$$

where $X_1, X_2, \ldots$ have a standard exponential distribution. Using Cramér's theorem, it is easy to see that $\{\nu_n\}$ satisfies the LDP with the non-convex good rate function

$$I(x) := \begin{cases} x - 1 - \log x & \text{if } x \in (0, \frac{5}{4}\log 5]; \\ \frac{x}{5} - 1 - \log\frac{x}{5} & \text{if } x > \frac{5}{4}\log 5; \\ \infty & \text{otherwise.} \end{cases}$$

In order to avoid simulation, one could try to compute $\nu_n(A)$ by calculating $\nu_n^1(A)$ and $\nu_n^5(A)$ numerically, where ν_n^1 is the law of $\frac{1}{n}\sum_{i=1}^n X_i$ and ν_n^5 is the law of $\frac{5}{n}\sum_{i=1}^n X_i$. Indeed, both probabilities are readily expressed in terms of (incomplete) Gamma functions. However, numerical problems arise already for moderate values of n, since one has to divide an incomplete Gamma function by $(n-1)!$.

Therefore, it is natural to consider estimation of $\nu_n(A)$ using simulation techiniques, and there are several possibilities. Since α is known, it suffices to simulate $\nu_n^1(A)$ and $\nu_n^5(A)$ separately. However, $\nu_n^1(A)$ cannot be simulated efficiently by twisting exponentially; refer to Glasserman and Wang [19] for related examples. To see that problems arise, we apply Corollary 1 for the simulation of $\nu_n^1(A)$, i.e., with the rate function $I^1(x) := x - 1 - \log x$. First note that $\inf_{x\in A} I^1(x) = I^1(1/10)$; one can readily check that the twist $\xi = -9$ is the only candidate twist for asymptotic efficiency. However, this twist cannot be used for simulation since the condition in Corollary 1 is violated: although $\inf_{x\in\mathbb{R}} I^1(x) + 9x = I^1(1/10) + 9/10$, one also has $\inf_{x\in A} I^1(x) - 9x < I^1(1/10) - 9/10$. In the setting of this example, this can easily be overcome at some additional computational costs: one can simulate $\nu_n^1((10,\infty))$ and $\nu_n^1([0, 1/10))$ separately. These probabilities and $\nu_n^5(A)$ can be simulated efficiently with an exponential twist, so that a reliable estimate of $\nu_n(A)$ is found by simulating *three* different probabilities.

However, it is more efficient to take a direct approach in this case: the reader easily checks that the twist $\xi = 1/10$ is asymptotically efficient as a direct consequence of Corollary 1 (applied to the rate function I). For the direct approach to be more efficient than the 'cut approach', it is essential that one can easily sample from the ξ-twisted distribution.

The ξ-twisted measure is λ_n, where for Borel sets A,

$$\lambda_n(A) = \frac{\alpha \int_A \exp(n\xi x)\, \nu_n^1(dx) + (1-\alpha) \int_A \exp(n\xi x)\, \nu_n^5(dx)}{\alpha \int \exp(n\xi y)\, \nu_n^1(dy) + (1-\alpha) \int \exp(n\xi y)\, \nu_n^5(dy)} \tag{5.1}$$

$$= \overline{\alpha}_n \frac{\int_A \exp(nx/10)\, \nu_n^1(dx)}{\left(\frac{10}{9}\right)^n} + (1-\overline{\alpha}_n) \frac{\int_A \exp(nx/10)\, \nu_n^5(dx)}{2^n}, \tag{5.2}$$

with

$$\overline{\alpha}_n = \frac{\alpha}{\alpha + (1-\alpha)\left(\frac{9}{5}\right)^n}.$$

Representation (5.2) explains how one can sample from λ_n: with probability $\overline{\alpha}_n$, one draws from an Erlang$(n, \frac{10}{9}n)$ distribution, and with probability $1-\overline{\alpha}_n$, one draws from an Erlang$(n, 10n)$ distribution. Observe that $\overline{\alpha}_n \to 0$; this is quite natural as the mean of the twisted distribution then tends to $10 = \arg\inf_{x\in A} I(x)$. The likelihood ratio can be written as follows, cf. (5.1):

$$\frac{d\lambda_n}{d\nu_n}(x) = \frac{\exp(nx/10)}{\alpha\left(\frac{10}{9}\right)^n + (1-\alpha)2^n}.$$

Therefore, in the direct approach only one probability is simulated instead of three as in the 'cut approach'. On the other hand, one has to draw from the distribution $(\overline{\alpha}_n, 1-\overline{\alpha}_n)$.

6. Discussion

In case any exponential twist for estimating $\nu(A)$ is asymptotically inefficient, there are a number of alternatives. First of all, it may be possible to write the rare event A as a union of $m < \infty$ disjoint rare events $A_1, \ldots, A_m$, for which the probabilities can be estimated efficiently by an exponential twist; see also Section 5. The sum of these probabilities is then an asymptotically efficient estimator for $\nu(A)$. In many applications, however, A cannot be written in that form. To overcome this, one can approximate $\nu(A)$ by $\nu(\bigcup_{i=1}^m A_i)$ for suitably chosen $A_1, \ldots, A_m$ and bound the error in some sense, as in [5] and [11]. A variant of this approach is based on mixing relevant exponential twists; details can be found in Sadowsky and Bucklew [26]. In a hitting probability framework, Collamore [8] uses related ideas to find an estimator that is arbitrarily 'close' to asymptotic efficiency.

Another possibility to deal with asymptotically inefficient exponential twists is the recent adaptive approach to importance sampling described by Dupuis and Wang [14]. Although the authors illustrate this approach in an setting based on Cramér's Theorem, they claim it is useful in a more general setting. This dynamic exponential twisting contrasts with the approach taken in this paper, as we consider a fixed exponential twist.

Although our definition of asymptotic efficiency is mathematically convenient, several other criteria for discriminating between estimators have been proposed. Notably, the amount of time (or work) required to generate one simulation replication was not taken into account in our definition of asymptotic efficiency. Glynn and Whitt [20] elaborate on a definition in which this is incorporated.

References

1. S. Asmussen, *Risk theory in a Markovian environment*, Scand. Actuar. J., (1989), pp. 69–100.

2. ——, *Applied probability and queues*, Springer-Verlag, New York, second ed., 2003.

3. S. Asmussen and R. Y. Rubinstein, *Steady state rare events simulation in queueing models and its complexity properties*, in Advances in queueing, J. Dshalalow, ed., CRC, Boca Raton, FL, 1995, pp. 429–461.

4. J. D. Biggins, *Large deviations for mixtures*. Preprint, 2003.

5. N. K. Boots and M. Mandjes, *Fast simulation of a queue fed by a superposition of many (heavy-tailed) sources*, Probab. Engrg. Inform. Sci., 16 (2002), pp. 205–232.

6. J. A. Bucklew, P. Ney, and J. S. Sadowsky, *Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains*, J. Appl. Probab., 27 (1990), pp. 44–59.

7. N. R. Chaganty, *Large deviations for joint distributions and statistical applications*, Sankhyā Ser. A, 59 (1997), pp. 147–166.

8. J. F. Collamore, *Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors*, Ann. Appl. Probab., 12 (2002), pp. 382–421.

9. A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, New York, second ed., 1998.

10. J.-D. Deuschel and D. W. Stroock, *Large deviations*, Academic Press Inc., Boston, MA, 1989.

11. A. B. Dieker and M. Mandjes, *Fast simulation of overflow probabilities in a queue*

with Gaussian input, Tech. Rep. PNA-E0404, CWI, the Netherlands, 2004.

12. I. H. Dinwoodie and S. L. Zabell, *Large deviations for exchangeable random vectors*, Ann. Probab., 20 (1992), pp. 1147–1166.
13. ——, *Large deviations for sequences of mixtures*, in Statistics and Probability: A Raghu Raj Bahadur Festschrift, J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy, and B. L. S. Prakasa Roa, eds., Wiley Eastern, New Delhi, 1993, pp. 171–190.
14. P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, tech. rep., Lefschetz Center for Dynamical Systems, Brown University, 2002.
15. J. Feng and T. Kurtz, *Large deviations for stochastic processes*. Preprint, 2000.
16. P. Glasserman, P. Heidelberger, and P. Shahabuddin, *Asymptotically optimal importance sampling and stratification for pricing path-dependent options*, Math. Finance, 9 (1999), pp. 117–152.
17. P. Glasserman and S. Kou, *Analysis of an importance sampling estimator for tandem queues*, ACM Trans. Modeling Comp. Simulation, 4 (1995), pp. 22–42.
18. P. Glasserman and J. Li, *Importance sampling for portfolio credit risk*. Preprint, 2003.
19. P. Glasserman and Y. Wang, *Counterexamples in importance sampling for large deviations probabilities*, Ann. Appl. Probab., 7 (1997), pp. 731–746.
20. P. W. Glynn and W. Whitt, *The asymptotic efficiency of simulation estimators*, Oper. Res., 40 (1992), pp. 505–520.
21. P. Heidelberger, *Fast simulation of rare events in queueing and reliability models*, ACM Trans. Modeling Comp. Simulation, 5 (1995), pp. 43–85.
22. T. Lehtonen and H. Nyrhinen, *On asymptotically efficient simulation of ruin probabilities in a Markovian environment*, Scand. Actuar. J., (1992), pp. 60–75.
23. ——, *Simulating level-crossing probabilities by importance sampling*, Adv. in Appl. Probab., 24 (1992), pp. 858–874.
24. J. S. Sadowsky, *On the optimality and stability of exponential twisting in Monte Carlo estimation*, IEEE Trans. Inform. Theory, 39 (1993), pp. 119–128.
25. ——, *On Monte Carlo estimation of large deviations probabilities*, Ann. Appl. Probab., 6 (1996), pp. 399–422.
26. J. S. Sadowsky and J. A. Bucklew, *On large deviations theory and asymptotically efficient Monte Carlo estimation*, IEEE Trans. Inform. Theory, 36 (1990), pp. 579–588.
27. D. Siegmund, *Importance sampling in the Monte Carlo study of sequential tests*, Ann. Statist., 4 (1976), pp. 673–684.
28. R. Szechtman and P. W. Glynn, *Rare-event simulation for infinite server queues*, in Proceedings of the 2002 Winter Simulation Conference, dec 2002.